METHOD AND SYSTEM FOR FREQUENCY SPECTRUM RESOURCE ALLOCATION

Background of the Invention

1. Field of the Invention

The invention relates generally to wireless communication systems. More particularly, this invention relates to optimizing the allocation of the frequency spectrum among several stations of a wireless communication system.

2. Description of the Related Art

Wireless communication systems provide for the transmission and reception of voice, data, and video information among multiple stations (e.g., remote units) over radio frequency (RF) channels. The RF spectrum is limited by its very nature and, consequently, only a small portion of the spectrum can be assigned to a particular industry. Hence, in an industry such as the satellite communication or cellular telephone industries, designers are continuously challenged to efficiently allocate the limited spectrum to allow as many remote units as possible to have access to the assigned frequency spectrum.

One way of satisfying the demands of this challenge includes implementing one or more modulation techniques. Some modulation techniques, such as time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA), have demonstrated an efficient spectrum utilization. Each of these access techniques is well known in the art and, hence, will not be described herein. Generally, each of these techniques provides a method of accessing a particular segment of the spectrum by multiple contending remote units (e.g., users). These techniques, however, do not account or adapt for variations in propagation conditions when allocating a particular segment of the spectrum to the multiple users. For example, in a satellite system that employs a TDMA technique, a user is typically allocated a particular periodic timeslot (on a predetermined frequency) during which the user may communicate with a hub station. To allow multiple users to communicate with the hub station, multiple non-overlapping timeslots are allocated to multiple users,

respectively. In nearly every wireless system, however, signal propagation may be subject to unpredictable degradation over one or more intervals of time. Generally, there are several physical phenomena which introduce degradation in the wireless medium. For example, in satellite communication systems, signal degradation may be caused by weather conditions (e.g., rain storms) or environmental interference. In land based communication systems, signal degradation may be caused by physical phenomena, such as multipath propagation and varying distance between the transmitter and receiver. Such signal degradation adversely affects channel performance for some users, but not necessarily for others.

Moreover, these sophisticated access techniques do not accommodate for or respond to changes in the utilization of the assigned spectrum among various users. For example, during a particular interval of time one user may be in need of transmitting an amount of information that, if transmitted with the current bandwidth, may take an unreasonable length of time. During the same interval of time, another user may not have such a need and be idle. This situation is particularly common in data communication networks, such as the Internet, where data is transmitted in bursts or packets (i.e., chunks of bits) between one communication station and another. The bursty characteristic of such networks renders conventional frequency spectrum utilization inefficient.

Therefore, there is a need in the industry to dynamically allocate frequency spectrum utilization in user demand and performance, so that all users have adequate access to the assigned spectrum.

Summary of the Invention

To overcome the above-mentioned limitation, the invention provides a method and system for optimizing frequency spectrum utilization. The invention provides a method of allocating at least a portion of the radio frequency (RF) spectrum among a plurality of RF transmitters. The method comprises monitoring aggregate demand of a group of transmitters within the plurality of RF transmitters. The group comprises at least one RF transmitter. The method further comprises determining, in response to the monitored demand, relative data congestion of the group of transmitters. The method

further comprises allocating at least a portion of the RF spectrum from a group having a least amount of congestion to at least one of the plurality of other RF transmitters.

The invention further provides a system for allocating at least a portion of the radio frequency (RF) spectrum among a plurality of RF transmitters. The system comprises a plurality of RF transmitters each configured to transmit data over a respective RF channel. The system further comprises a hub transceiver in communication with the plurality of RF transmitters. The hub transceiver is configured to monitor the aggregate demand of a group of the plurality RF transmitters. The group comprises at least one RF transmitter. The hub transceiver is further configured to re-allocate a portion of the RF spectrum from the group of RF transmitters having a smallest aggregate demand to at least one of the plurality of other RF transmitters

Brief Description of the Drawings

The above and other aspects, features, and advantages of the invention will be better understood by referring to the following detailed description, which should be read in conjunction with the accompanying drawings, in which:

Figure 1 is a block diagram of a typical satellite communication system in which the invention may be implemented.

Figure 2 is a block diagram of a wireless communication system comprising a base station and multiple remote units in accordance with the invention.

Figure 3 is a flowchart describing the process of determining whether to allocate the frequency spectrum among two or more groups of the wireless communication system of Figure 2.

Figure 4 is a flowchart describing the process of determining the aggregate demand of one or more groups of the wireless communication system of Figure 2.

Figure 5 is a flowchart describing the process of determining congestion and re-allocation the frequency spectrum among two or more groups of remote units of the wireless communication system of Figure 2.

Figure 6 is a table showing exemplary groups of the remote units of the wireless communication system of Figure 2.

Figure 7 is a table showing an exemplary change in the groups of wireless communication system of Figure 2.

Figure 8 is a graphical representation of one embodiment of the process of re-allocating the frequency spectrum among the remote units as a function of frequency and time.

Figure 9 is a graphical representation of three regions of quality of service operating regions for a remote unit.

Figure 10 is a flowchart describing the process of dynamically scheduling remote unit communications in accordance with another embodiment of the invention.

Figure 11 is a graphical representation of the exemplifying results of the process of scheduling remote unit communications as a function of frequency and time.

Detailed Description of the Invention

The following description is not to be taken in a limiting sense, but is made merely for the purpose of describing the general principles of the invention. Like components are identified with like component numbers throughout the following description. The scope of the invention should be determined with reference to the claims.

Figure 1 is a block diagram showing an exemplary system 150 in which the invention may be embodied. The system 150 provides high-speed, reliable Internet communication service over a satellite link.

In particular, the system 150 comprises on or more content servers 100 that are coupled to the Internet 102, which is in turn coupled to a hub station 104. The hub station 104 is configured to request and receive digital data from the content servers 100. The hub station 104 also communicates via a satellite 106 with a plurality of remote units 108A – 108N. For example, the hub station 104 transmits signals over a forward uplink 110 to the satellite 106. The satellite 106 receives the signals from the forward uplink 110 and re-transmits them on a forward downlink 112. Together, the forward uplink 110 and the forward downlink 112 are referred to as the forward link. The remote units 108A – 108N monitor one or more channels which comprise the

forward link in order to receive remote-unit-specific and broadcast messages from the hub station 104.

In a similar manner, the remote units 108A – 108N transmit signals over a reverse uplink 114 to the satellite 106. The satellite 106 receives the signals from the reverse uplink 114 and re-transmits them on a reverse downlink 116. Together, the reverse uplink 114 and the reverse downlink 116 are referred to as the reverse link. The hub station 104 monitors one or more channels which comprise the reverse link in order to extract messages from the remote units 108A – 108N. For example, in one embodiment of the system 150, the reverse link carries multiple access channels as described in assignee's co-pending application entitled METHOD AND APPARATUS FOR MULTIPLE ACCESS IN A COMMUNICATION SYSTEM, Application Serial No. __/___,___, [Attorney Docket No. TACHYON.018CP2], filed concurrently herewith, the entirety of which is hereby incorporated by reference.

In one embodiment of the system 150, each remote unit 108A –108N is coupled to a plurality of system users. For example, in Figure 1, the remote unit 108A is shown as coupled to a local area network 116, which in turn is coupled to a group of user terminals 118A – 118N. The user terminals 118A - 118N may be one of many types of local area nodes such as a personal or network computer, a printer, digital meter reading equipment or the like. When a message is received over the forward link intended for one of the user terminals 118A – 118N, the remote unit 108A forwards it to the appropriate user terminal 118 over the local area network 116. Likewise, the user terminals 118A – 118N can transmit messages to the remote unit 108A over the local area network 116.

In one embodiment of the system 150, the remote units 108A – 108N provide Internet service to a plurality of users. For example, the user terminal 118A may be a personal computer that executes browser software in order to access the World Wide Web. When the browser receives a request for a web page or embedded object from the user, the user terminal 118A creates a request message according to well known techniques. The user terminal 118A forwards the request message over the local area network 116 to the remote unit 108A, also using well known techniques. Based upon the request message, the remote unit 108A creates and transmits a wireless link request

over a channel of the reverse uplink 114 and the reverse downlink 116. The hub station 104 receives the wireless link request over the reverse link. Based upon the wireless link request, the hub station 104 passes a request message to the appropriate content server 100 over the Internet 102.

In response to the request message, the content server 100 forwards the requested page or object to the hub station 104 over the Internet 102. The hub station 104 receives the requested page or object and creates a wireless link response. The hub station transmits the wireless link response over a channel of the forward uplink 110 and forward downlink 112. For example, in one embodiment of the system 150, the hub station 104 operates as described in assignee's co-pending application entitled TRANSMISSION OF TCP/IP DATA OVER A WIRELESS COMMUNICATION CHANNEL, Application Serial No. __/__,__ [Attorney Docket No. TACHYON.046A], which is filed concurrently herewith and the entirety of which is hereby incorporated by reference.

The remote unit 108A receives the wireless link response and forwards a corresponding response message to the user terminal 118A over the local area network 116. In one embodiment of the system 150, the process of retrieving a web page or object is executed as described in assignee's co-pending application entitled DISTRIBUTED SYSTEM AND METHOD FOR PREFETCHING OBJECTS, Application Serial No. 09/129,142 [Attorney Docket No. TACHYON.001A2], filed August 5, 1998, the entirety of which is hereby incorporated by reference. In this way, a bi-directional link between the user terminal 118A and the content servers 100 is established.

As noted above, the invention provides a method and system for optimizing frequency spectrum utilization in response to changes in demand of the remote units. There are several ways to assess the condition of a channel of a particular remote unit in a wireless system. One common way involves estimating the signal-to-noise ratio (SNR) of the signals received from the remote unit. SNR is the measure of the energy of the signal (usually expressed in decibels or dB), over a predetermined bandwidth and/or time interval, relative to the energy of the noise added to the signal. Generally, "noise" refers to the difference between the signal transmitted by one of the remote

units and the signal received by the hub station 104. The higher the SNR of a channel, the better is the channel performance.

Another common way of characterizing channel performance involves estimating the bit error rate (BER) of the channel. Simply stated, BER is expressed as the fraction of the number of bits received incorrectly over the total number of bits transmitted. BER is either expressed as a percentage or, more usually, as a ratio. In effect, BER is a measure of the probability of bit error in the channel. The lower the BER, the better is the channel performance.

Figure 2 shows a block diagram of a wireless communication system 200 comprising a hub station 210 and representative remote units 212, 214, 216, 232, 234, 252, and 254 in accordance with the invention. The system 200 may comprise a satellite-based wireless system (as shown in Figure 1), or any other wireless system (e.g., mobile telephone) having multiple remote units. The system 200 may apply TDMA, FDMA, any other access technique, or a combination of access techniques to implement the invention. The number of stations in the system 200 is only illustrative and, hence, the system 200 may comprise any desired number of hub and remote stations.

The remote units are categorized into two or more operational groups of remote units (sometimes referred to as "camps" of remote units) based on the assigned data rate of each remote unit. In one embodiment, the system 200 comprises three groups of remote units: Group 32, Group 64, and Group 128. Group 32 includes one or more remote units that operate at a data rate of 32 kbps, Group 64 includes one or more remote units that operate at a data rate of 64 kbps, and Group 128 includes one or more remote units that operate at a data rate of 128 kbps. Typically, the hub station 210 determines and communicates the assigned data rate to each remote unit. For example, the hub station 210 may assign a data rate of 32 kbps to the remote units 212, 214, and 216, thereby placing these remote units in Group 32. Similarly, the hub station 210 may assign a data rate of 64 kbps to the remote units 232 and 234, thereby placing these remote units in Group 64. Finally, the hub station may assign a data rate of 128 kbps to the remote units 252 and 254, thereby placing these remote units in Group 128.

The hub station 210 determines and assigns the data rate to each remote unit based on its respective channel condition. The channel condition indicates the ability of the channel to sustain an assigned data rate while still maintaining an acceptable signal performance (e.g., SNR). In one embodiment, the hub station 210 is configured to continuously, or at predetermined time intervals, monitor the channel performance based on signals received from each of the remote units. More particularly, the hub station 210 may measure the SNR over a predetermined time interval to assess channel performance of each remote unit. The hub station 210 compares the measured SNR to predetermined SNR threshold values. The SNR thresholds may comprise a low-threshold value (e.g., 8 dB) and a high-threshold value (e.g., 11 dB). Based on this comparison, the hub station 210 determines whether to change the currently assigned data rate for each remote unit and, consequently, whether to re-categorize the remote unit from one group to another.

For example, if the measured SNR of the signals received from the remote unit 232 falls within the low and high thresholds, the hub station 210 determines that the remote unit 232 is operating at an optimal data rate and, hence, no change of the assigned data rate is necessary. If the measured SNR is above the high-threshold value, the hub station 210 determines that the channel of the remote unit 232 can sustain a higher data rate. Accordingly, the hub station 210 may instruct the remote unit 232 to raise its data rate from 64 kbps to a higher data rate, e.g., 128 kbps. If, on the other hand, the measured SNR is below the low-threshold value, the hub station 210 determines that the channel utilization of the remote unit 232 is unacceptable and its presently assigned data rate should be reduced. Accordingly, the hub station 210 may instruct the remote unit 232 to decrease its data rate from 64 kbps to a lower data rate, e.g., 32 kbps. The hub station 210 may repeat this process to optimize channel utilization for all remote units. In one embodiment, the average transmit power at each remote unit is unaffected and maintained substantially fixed throughout this process.

In addition, the hub station 210 is configured to dynamically allocate portions of the assigned frequency spectrum to the remote units in response to changes in the demand of each remote unit. As used herein, the term "demand" refers to the amount of information (e.g., data expressed in bits) that a remote unit desires to exchange or

transmit at a particular instant of time. Typically, the system 200 uses a channel, such as a reservation channel, on which each remote unit periodically, or when requested, reports or transmits its current demand to the hub station 210. In one embodiment, the hub station 210 is configured to determine the collective demand of the remote units on a group-by-group basis (hereinafter "aggregate demand"). As will be discussed in greater detail below, based at least in part on the aggregate demand of each of Groups 32, 64, and 128, the hub station 210 determines the portion of the frequency spectrum to be allocated to each of the Groups 32, 64, and 128. By so doing, the hub station 210 continuously reduces congestion and transmission delay of and optimizes the frequency utilization among the groups of remote units.

In one embodiment, it is desirable to check the quality of service (QoS) assigned to each remote unit before determining the aggregate demand of each of the Groups 32, 64, and 128. Generally, QoS may specify a nominal guaranteed throughput level (e.g., amount of data in bits) or data rate (expressed in kbps) for each remote unit. The QoS is typically assigned to each remote unit pursuant to a subscription agreement between the remote unit and the service provider, e.g., owner of the hub station 210. As used herein, the term "QoS" refers to any one or more criteria that a hub station 210 may use to classify the quality of performance committed or provided to a remote unit.

In general, the hub station 210 may use any communication parameter to allocate one or more portions of the frequency spectrum among the remote units. The communication parameter may include the aggregate demand of a group of remote units, individual demand of a single remote unit, quality of service, channel performance (e.g., SNR or BER measurements), number of remote units in a group, propagation paths (e.g., distance, terrain, etc.), any other parameter that affects performance of the wireless communication system 200, or any combination of these parameters. As further discussed below, based on the communication parameter, the hub station 210 determines the current or anticipated state of performance of the group of remote units (or a single remote unit) to allocate one or more portions of the frequency spectrum.

Figure 3 is a flowchart describing the process of determining whether to re-allocate the frequency spectrum among two or more groups of remote units of the

system of Figure 2. As noted above, in one embodiment, the remote units are categorized or distributed among Groups 32, 64, and 128. The process begins at block 300 when the system 200 initiates an algorithm to check channel performance for each remote unit. For example, the algorithm may be implemented using any microprocessor-based instructions, such as conventional firmware, programmed in, or on a device within fast access of the hub station 210. At block 310, the system 210 monitors a channel by listening for signals received from a first remote unit (e.g., the remote unit 232). In one embodiment, each remote unit may transmit signals to the hub station 210 over a predetermined or other available channel during periodic time intervals. The hub station 210 measures the energy of the signal and noise components of the signals arriving from the remote unit 232. As described above, the hub station calculates the SNR over a predetermined time interval (e.g., 100 milliseconds) for the remote unit 232.

At the decision block 320, the hub station 210 determines whether to change currently assigned data rate of the remote unit 232 based on the measured SNR. As noted above, the hub station 210 is programmed with low (e.g., 8 dB) and high (e.g., 11 dB) threshold values to compare the measured SNR. The range between the low and high thresholds represents sufficient channel performance for the currently assigned data rate. Accordingly, if the measured SNR falls within the low and high thresholds, the process proceeds to block 330 where the hub station 210 maintains the currently assigned data rate for the remote unit 232. In this case, the process continues to block 370 where the hub station 210 determines if all of the remote units in the group are checked, as described below.

The range of SNR below the low threshold value represents an undesirable channel performance where the noise level is relatively high for the currently assigned data rate. Hence, if the measured SNR falls below the low threshold value, the process proceeds to block 340 where the hub station 210 instructs the remote unit 232 to reduce its currently assigned data rate from 64 kbps to a lower data rate, e.g., 32 kbps. Thus, in such an event, the hub station 210 re-categorizes the remote unit 232 from Group 64 to Group 32. On the other hand, the range of SNR above the high threshold value represents an inefficient use of the channel where the noise level is relatively low for the

currently assigned data rate. Hence, in block 320 if the measured SNR falls above the high threshold value, the process proceeds to block 350 where the hub station 210 instructs the remote unit 232 to increase its currently assigned data rate from 64 kbps to a higher data rate, e.g., 128 kbps. Thus, in such an event, the hub station 210 re-categorizes the remote unit 232 from Group 64 to Group 128.

At block 360, the hub station 210 collects one or more signals representative of the demand of the remote unit 232 over the reservation channel. The hub station 210 stores the demand in an accessible memory (not shown) for later retrieval. The timing of collection of the demand signal may not be material to the invention and, hence, may be performed before, during, or after the SNR measurement of each remote unit. For instance, the hub station 210 may collect and save the demands of all of the remote units before initiating the process of Figure 3. At block 370, the hub station 210 determines if the demands from all of the remote units are obtained. If the demand of more remote units is still needed, the process may return to block 310 to measure SNR of the remaining remote units and repeat the process described thus far. Alternatively, the process may return to block 360 to collect the demand of the remaining remote units over the reservation channel. In one embodiment, one or more of these steps are performed in parallel.

If, on the other hand, the demand of all of the remote units is collected, the hub station 210 determines in block 380 if one or more of the Groups 32, 64, 128 is relatively congested. This process is described in greater detail below with reference to Figure 4. If the hub station 210 determines that no congestion is detected, the process returns to block 310 to run the entire process again. Optionally, the process may be terminated at this stage and restarted at a later time. If, on the other hand, the hub station 210 determines that one or more of the Groups 32, 64, and 128 is congested, the process proceeds to re-allocate the frequency spectrum from the least congested (i.e., best state of performance) group to the other groups. Hence, at block 390, the hub station 210 reduces the portion of the allocated frequency spectrum of the least congested group, and increases the portion of the allocated frequency spectrum of the other groups. This process is described in greater detail below with reference to Figure 5. The process terminates at block 398.

Figure 4 is a flowchart describing the process executed in block 380 of Figure 3 of determining the aggregate demand of one or more groups. The process begins with block 400. As indicated above, the hub station 210 may be configured to determine relative congestion of each of the Groups 32, 64, and 128. At block 410, the hub station 210 monitors the demand of a remote unit over the reservation channel. As noted above, the demand represents the amount of data (expressed in bits) that the remote unit desires to exchange or transmit at a particular instant of time. At block 420, the hub station 210 qualifies the received demand by checking the QoS assigned to the remote unit. The hub station 210 usually stores, or has at least access to, the QoS of each remote unit operating within its coverage area. By qualifying the demand, the hub station 210 checks the QoS of the remote unit to determine whether the QoS permits allocation of resources to satisfy the entire demand or not. Pursuant to the decision block 430, if the QoS permits satisfying the entire requested demand of the remote unit, then in block 440, the hub station 210 takes the entire demand into consideration when assessing the aggregate demand of one of the Groups 32, 64, and 128. If, on the other hand, the QoS does not permit the requested demand, the hub station 210, in block 450, determines a reduced demand (i.e., downsizes the demand) for the remote unit, and considers the reduced demand when assessing the aggregate demand of the group.

For example, the remote unit 212 may be assigned a QoS criterion that permits it to exchange up to 32 kilobits of data per second, thereby yielding an average amount of data of 1.92 (i.e., about 2) Megabits every minute. If at 12:00:00 hours, the remote unit 212 transmits 1 Megabit, the hub station 210 checks the QoS of the remote unit 212 and determines that up to about 2 Megabits is allowed. Hence, at 12:00 hours, the hub station 210 considers the entire 1 Megabit for assessing aggregate congestion for Group 32. If, however, at 12:00:30 hours (i.e., 30 seconds later), the remote unit 212 requests a demand to transmit 2 Megabits, the hub station 210 determines that, based on the QoS of the remote unit 212, a demand of only about 1 Megabit is permitted for the balance of the one minute interval, i.e., during 12:00:00-12:00:01. Accordingly, for the purpose of assessing aggregate congestion for Group 32 at 12:00:30, the hub station 210 downsizes the demand from 2 Megabits to about 1 Megabit.

For each group, the hub station 210 computes the aggregate demand of the group based on the collective demand of all of the remote units within the group. Hence, at the decision block 460, the hub station 210 checks to see if the demand was polled from all of the remote units of the group. If the demand of more remote units remains to be polled, the process returns to block 410. If, on the other hand, the hub station 210 determines that the demand was polled from all of the remote units of the group, the process proceeds to block 470. To determine the aggregate demand of a single group, the hub station 210, in block 470, adds up the demands and/or reduced demands of all of the remote units of the group. The aggregate demand represents an estimate of the (average) length of a queue of bits for the group. The hub station 210 may repeat the process for all of the Groups 32, 64, and 128, and stores the aggregate demand of all groups in its memory to perform congestion analysis. The process terminates in block 480.

There are several ways to analyze congestion for each group of remote units. In one embodiment, the hub station 210 determines the congestion of each group relative to a least congested group. Figure 5 is a flowchart describing the process of determining congestion and re-allocation of the frequency spectrum among two or more groups of remote units. The process begins at block 500. At block 510, the hub station 210 identifies the least congested group, which is typically the group that has the smallest queue length. Once the least congested group is identified, the hub station 210, in block 520, compares the queue length of the other groups with the queue length of the least congested group. By this comparison, the hub station may compute the percentage of excess bits by dividing the queue length of a group by the queue length of the least congested group. The percentage of excess bits represents the extent of congestion in one group relative to the least congested group. For example, the average queue length of each of Groups 32, 64, and 128 may be 100, 300, and 250 Megabits, respectively. In this example, Group 32 having a queue length of 100 Megabits represents the least congested group. The percentage of excess bits for Group 64 is 300% (or 300/100), and for Group 128 is 250% (or 250/100). As shown by this example, the percentage of excess bits is a number that may not be smaller than 100%,

because the queue length of any group is always greater than (or equal to) the queue length of the least congested group.

At block 530, the hub station 210 determines whether, based on the relative congestion of the groups, it is necessary to re-allocate a portion of the frequency spectrum from the least congested group to the other groups. In one embodiment, the hub station 210 bases its determination on the percentage of excess bits. For example, the hub station 210 may be configured to re-allocate the frequency spectrum only for the groups having a percentage of excess bits of 200% or greater. Hence, based on the above numerical example, the hub station 210 may remove portions of the frequency spectrum from Group 32 and assign it to Groups 64 and 128. Accordingly, if the re-allocation of the frequency spectrum is warranted to relieve congestion, the process proceeds to block 540. If, on the other hand, the re-allocation of the frequency spectrum is not warranted, the process terminates at block 560.

At block 540, the hub station 210 determines the amount of frequency spectrum (i.e., size of bandwidth) to be allocated from the least congested group to the other groups. Bandwidth commonly refers to the amount of data that can be transmitted in a given period over a transmission channel such as a radio transmitter. Typically, bandwidth is expressed in cycles per second (hertz or Hz) or bits per second (bps). It is desirable to minimize the amount bandwidth to be re-allocated from the least congested group to the other groups. By minimizing the amount of re-allocated bandwidth, the probability of queue oscillation and, hence, system instability is reduced. Queue oscillation commonly refers to the transfer of congestion between a least congested group and other groups back and forth, i.e., in an oscillating manner.

To minimize queue oscillation, it is desirable to re-allocate the bandwidth in a stepwise fashion from the least congested group to the other groups. In one embodiment, using the stepwise fashion, the hub station 210 may re-allocate bandwidth in unit increments to the higher congested groups. For example, using the above numerical example, the hub station 210 may re-allocate a bandwidth of 64 kbps from Group 32 to Group 64, and a bandwidth of 128 kbps from Group 32 to Group 128. The purpose of re-allocating the bandwidth is to relieve congestion in the groups having greater congestion. Accordingly, at block 550, the hub station 210 re-allocates portions

of the frequency spectrum from the least congested group to the other groups. The re-allocation process terminates at block 560.

In one embodiment, the hub station 210 continuously, or during predetermined time intervals, repeats the process of Figure 5. The relief of congestion in the other groups may increase likelihood of congestion in the least congested group. However, the ability of the hub station 210 to continuously monitor group congestion, and re-distribute the assigned frequency spectrum among the groups of remote units, reduces the likelihood of congestion in a single group. Moreover, the continuous monitoring and re-allocation of the frequency spectrum optimizes frequency utilization among the remote units.

Figure 6 is a table showing exemplary groups of the remote units of Figure 2. As noted above, the hub station 210 assigns each remote unit to a camp or group based on the data rate assigned to each remote unit. In the table 600, the hub station 210 assigns a data rate of 32 kbps to remote units 212-224 and 244-246 and, thus, these remote units belong to Group 32. Similarly, the hub station 210 assigns a data rate of 64 kbps to remote units 232-242 and, thus, these remote units belong to Group 64. Finally, the hub station 210 assigns a data rate of 128 kbps to remote units 252-270 and, thus, these remote units belong to Group 128. As noted above, the data rate is generally assigned to each remote unit based on its channel performance, e.g., the measured SNR of the signals transmitted from each remote unit and received at the hub station 210. As explained above, if the SNR falls within an optimal range, the currently assigned data rate of the remote unit is maintained. If the SNR falls below a low or above a high threshold value, the data rate of the remote unit is reduced or increased accordingly. The hub station 210 maintains the table 600 in memory, or within easy access, to keep track of and update each group of remote units.

Figure 7 is a table showing an exemplary change in the Groups 32, 64, and 128. In this embodiment, the table 700 shows that the remote units 244 and 246 no longer belong to Group 32, but now belong to Group 64. Typically, a change in the grouping of the remote units 244 and 246 indicates that the measured SNR of the channel of each of the remote units 244 and 246 falls above the high threshold value. In that case, the hub station 210 instructs the remote units 244 and 246 to increase their respective data

rates from 32 to 64 kbps. Accordingly, the hub station 210 updates the table 600 to the table 700, which shows that the remote units 244 and 246 belong to Group 64.

Figure 8 is a graphical representation of the process of re-allocating the frequency spectrum among the remote units as a function of frequency and time. The graph 800 includes a vertical axis that represents the portions of the frequency spectrum (e.g., bandwidth) assigned to each group. More particularly, the graph 800 shows that the bandwidth 832 is assigned to Group 32, bandwidth 864 is assigned to Group 64, and bandwidth 828 is assigned to Group 128. The graph 800 also includes a horizontal axis that represents the time domain T. Beginning at T=0, the graph 800 shows that the time interval during which each remote unit may communicate is represented by a box (or timeslot) marked by the remote unit number.

For example, during the time interval 0-t_3, the graph 800 shows that the remote unit 212 is allocated a timeslot 212 and carrier frequency F_8, and is operating in Group 32 at a data rate of 32 kbps. During the same time interval 0-t_3, the graph 800 shows that the remote unit 214 is allocated the timeslot 214 and carrier frequency F_7, and is operating in Group 32 at a data rate of 32 kbps. During the time interval 0-t_2, the graph 800 shows that the unit 232 is allocated a timeslot 232 and carrier frequency F_9, and is operating in Group 64 at a data rate of 64 kbps. During the time interval 0-t_1, the graph 800 shows that the remote unit 252 is allocated a timeslot 252 and carrier frequency F_{10}, and is operating in Group 128 at a data rate of 128 kbps.

In this embodiment, it can be seen that the duration of the timeslot for the remote units of Group 32 is twice as long as the timeslot for the remote units of Group 64, and four times as long as the timeslot for the remote units of Group 128. The relationship between the duration of the timeslots among the various groups is typically a function of the assigned data rate. For example, because the data rate of 64 kbps is twice the data rate of 32 kbps, it is expected that the duration of the timeslot of Group 64 will be half the duration of the timeslot of Group 32. This timeslot/frequency structure simplifies the implementation of TDMA and FDMA systems having various operating data rates. Finally, it can also be seen that, in all of the groups, each remote unit does not occupy more than a single timeslot concurrently. The occupation of a single timeslot simplifies the operation of single-channel transceiver systems. Once the

portion of the frequency spectrum for each group is determined, the hub station 210 may assign one or more timeslot/frequency to a particular remote unit (within a group) using any standard implemented in the hub station 210. Additional details concerning the transmission of channel assignment information to a plurality of remote units are disclosed in assignee's co-pending application entitled SYSTEM AND METHOD FOR EFFICIENT CHANNEL ASSIGNMENT, Application Serial No. ________ [Attorney Docket No. TACHYON.043A], filed concurrently herewith, the entirety of which is hereby incorporated by reference. The invention is not limited to only such systems, but may be implemented using any timeslot/frequency structure that is compatible with the characteristics of the invention.

The graph 800 illustrates an exemplary change in respective bandwidth among the groups in response to the hub station's decision to re-allocate the assigned frequency spectrum. As shown in Figure 8, at time $T = t_4$, the hub station 210 changes the frequency spectrum allocation among the groups of remote units. More particularly, the graph 800 shows that each of the bandwidths 828 and 864 is doubled in size, and the bandwidth 832 is reduced accordingly. Hence, instead of availing only a single timeslot to Group 128 before $T = t_4$, two concurrent timeslots are available to the remote units of Group 128 after $T = t_4$. For example, as of time $T = t_4$, it can be seen that the remote units 270 and 268 are concurrently communicating at an assigned data rate of 128 kbps (Bandwidth 828). Similarly, instead of availing only a single timeslot to Group 64 before $T = t_4$, two concurrent timeslots are available to the remote units of Group 64 after $T = t_4$. For example, as of time $T = t_4$, it can be seen that the remote units 240 and 242 are concurrently communicating at an assigned data rate of 64 kbps (Bandwidth 864). On the other hand, instead of availing eight concurrent timeslots to Group 32 before $T = t_4$, only two concurrent timeslots remain available to the remote units of Group 32 after $T = t_4$. This illustration shows that, in response to the relative congestion of each of the Groups 64 and 128, the hub station 210 has determined that such congestion warrants frequency re-allocation from the least congested Group 32 to Groups 64 and 128, as explained in detail above.

Furthermore, the graph 800 illustrates an exemplary change in the data rate of one or more remote units. As shown in Figure 8, it can be seen that before $T = t_5$, each

of the remote units 244 and 246 was operating in Group 32 (Bandwidth 832) at a data rate of 32 kbps, as shown by timeslots 244 and 246 of Group 32. After time $T = t_5$, however, the remote units 244 and 246 are operating in Group 64 (Bandwidth 864) at a data rate of 64 kbps, as shown by timeslots 244 and 246 of Group 64. Hence, the graph 800 demonstrates that, sometime during the interval t_4 - t_5, the hub station 210 determined to change the assigned data rate of remote units 244 and 246 from 32 to 64 kbps. As explained in detail above, the hub station 210 bases its determination on the measured SNR for the channel of each of the remote units 244 and 246. In this example, the SNR falls above a high threshold value (e.g., 11 dB), thereby warranting an increase in data rate. Accordingly, the hub station 210 instructs the remote units 244 and 246 to raise their respective data rates.

In another embodiment of the invention, the reverse link resources are not pre-assigned to particular camps of remote units. Figure 9 is an exemplary graphical representation of the three quality of service operating regions for a particular remote unit, e.g., remote unit 212 (see Figure 2) which operates in such an environment. As noted above, the QoS is typically allocated to each remote unit pursuant to a subscription agreement between the remote unit and the service provider. Independent of the assigned data rate, the QoS specifies an allocated average data rate. While the assigned data rate specifies the rate at which the remote unit is capable of transmitting information over the channel when the remote unit is allocated a resource, the allocated average data rate reflects the average data rate over some extended period which the remote unit has, for example, purchased from the service provider. For example, if a remote unit has an assigned data rate of 256 kbps and an allocated average data rate of 32 kbps, although the remote unit transmits in bursts at a rate of 256 kbps, the bursts are dispersed in time by idle periods which reduce the average data transfer rate of the remote unit to about 32 kbps. In other words, the average duty cycle of this remote unit's transmission is at most about one-to-eight.

Figure 9 shows a vertical axis 402 representing a range of current average data rates for the remote unit 212. Pursuant to its agreement, the remote unit 212 has subscribed for an allocated average data rate 404 (e.g., 32 kbps). Average data rates below this value are represented by an IN region 406. In one embodiment, it may be

desirable to allow the remote unit 212 to exceed its allocated average data rate 404, and permit operation in an OUT region 414. The OUT region 414 represents a range of average data rates at which the remote unit 212 may operate above its allocated average data rate 404. Hence, the OUT region 414 represents average data rates ranging from the allocated average data rate 404 to a maximum average data rate 408 (e.g., 48 kbps). As further shown in Figure 9, a HARD DROP region 412 represents average data rates above the maximum average data rate 408.

In one embodiment, the allocated average data rate 404 is associated with a particular remote unit in accordance with a subscription agreement between the remote unit operator and the owner or operator of the hub station. For example, a service provider may wish to reduce the operating costs associated with providing Internet services by purchasing a relatively low allocated average data rate 404. As the number of subscribers and the demand on the system increases, the service provider may purchase a higher allocated average data rate 404, presumably at a greater cost.

The quality of service levels 404 associated with remote units are stored by hub station. In one embodiment, the hub station includes tables that store a remote unit identifier and an associated allocated average data rate 404. In one embodiment, the tables are updated by the hub station operator when subscription information is added or modified.

Each hub station stores a range parameter that is used to define the data rate by which a transmission from a remote unit can exceed the allocated average data rate 404. The range parameter defines the size of the OUT region 414 by providing the value for the maximum average data rate 408. The range parameter may be selected based on typical system use, the capacity of the hub station, and other factors. The use of the maximum average data rate artificially limits the average data rate of a remote unit even if system resources are available, thus encouraging the purchase of a higher allocated average data rate. In other embodiments, the same mechanisms may be employed to limit the maximum average data rate pursuant to other motives.

In this embodiment, the invention provides a method and system of scheduling remote unit communications of the system 200 within the available communication resources. As noted above, the hub station 210 may continuously receive demand of

each of the remote units over the reservation channel. In this embodiment, the hub station 210 arranges each arriving demand in a queue on a first in first out (FIFO) basis.

In one embodiment, the hub station 210 categorizes or classifies each remote unit demand based, at least in part, on the current average data rate for the remote unit over some previous period of time. As indicated above, the hub station 210 may compute a current average data rate based on a moving average over a predetermined time interval (e.g., 10, 30, 60 seconds, or other desired interval). The moving average is determined by dividing the amount of data transmitted during a predetermined past time interval by the predetermined time interval.

For example, assume that a remote unit has an allocated average data rate of 48 kbps, a maximum average data rate of 60 kbps and that the hub station uses a predetermined time interval of 60 seconds for determining the remote unit's average data rate. Further, assume that, after a long period of idleness, at 12:00:01, the remote unit 212 completes a transfer of 1 Megabit of data. Hence, up to 12:00:02, the current average data rate of the remote unit is about 17 kbps (i.e., 1 Megabit/60 seconds), which places the remote unit 212 in the IN region 406. At 12:00:30 hours, the remote unit 212 completes the transfer of 2 Megabits of data. In view of the 1 and 2 Megabit transfers, the current average data rate of the remote unit 212 at time 12:00:31 is 50 kbps (3 Megabits/60 seconds), which places the remote unit's 212 operating point in the OUT region 414. Finally, if at 12:00:45, the remote unit 212 completes the transfers of 3 Megabits of data, the current average data rate of the remote unit 212 at time 12:00:46 is about 100 kbps (i.e., 6 Megabits/60 seconds), which places the remote unit's operating point in the DROP HARD region 412. If, as time continues to pass, the remote unit 212 does not transfer any more data, the remote unit's current average data rate eventually falls through the OUT region 414 and into the IN region 406.

Figure 10 is a flowchart describing a second embodiment of the process of dynamically scheduling remote unit communications. The process begins at start block 804. As noted above, in one embodiment, the hub station 210 receives demand requests from each remote unit that desires to communicate data over the system 200 (see Figure 2) and places a corresponding entry in a FIFO queue. At block 808, the hub

station 210 determines the current average data of the first remote unit corresponding to the first entry in the FIFO queue, for example, as just described.

At block 812, the hub station 210 determines if the current average data rate of the remote unit 212 classifies it as operating in the HARD DROP region 412 (see Figure 9). If, based on the amount of data transmitted by this remote unit over the predetermined interval (e.g., past 60 seconds), the remote unit 212 is operating in the HARD DROP region 412, the process proceeds to block 816 where the hub station 210 places the current demand entry to the end in the FIFO queue. By delaying satisfaction of the demand to a later time, the hub station 210 declines to grant a bandwidth/timeslot to the remote unit 212 at this time thereby reducing the current average data rate of the remote unit moving forward in time. In another embodiment, the demand entry is removed from queue and is not replaced with in the queue.

If, on the other hand, the remote unit 212 is not operating in the HARD DROP region 412 during the predetermined interval, the process proceeds to block 820 where the hub station 210 determines if the remote unit 212 is operating in the OUT region 414.

If, based on its current average data rate over the predetermined interval, the remote unit 212 is operating in the OUT region 414, the process continues to block 824 where the hub station 210 performs the OUT version of a pair of algorithms, such as Random Early Drop (RED) with In/Out bit (RIO). In one embodiment, the RED and RIO algorithms are executed by a gateway within the hub station. Generally, a RED algorithm computes the average queue length and, when the average queue length exceeds a certain dropping threshold, the gateway begins to randomly drop demand requests with a certain probability, where the exact probability is a function of the queue length at the hub stations.

If, based on its current average data rate over the predetermined interval, the remote unit 212 is operating in the IN region 406, the process continues to block 828 in which a second Random Early Drop (RED) algorithm is performed. Typically, the dropping threshold reflects a longer queue length for the IN packets than the OUT packets and the probability of dropping an OUT packet is higher than or equal to the probability of dropping an IN packet over the entire range of queue lengths. For further

details on the RED and RIO algorithms and gateways, reference is made to Clark, D. and Fang, W., Explicit Allocation of Best Effort Packet Delivery Service, which is available via http://diffserv.lcs.mit.edu/Papers/exp-alloc-ddc-wf.pdf.

If the demand request is not passed (i.e., is dropped) by the RED algorithm in either block 824 or 828, the process returns to block 816 where the hub station 210 places the demand entry at the end of the FIFO queue or drops the request from the FIFO queue. If, on the other hand, the demand request of the remote unit 212 is passed by the RED algorithm in either block 824 or 828, the process continues to block 830.

In block 830, the hub station 210 schedules the remote unit communication. More particularly, to schedule the remote unit communication, the hub station 210 determines the bandwidth that is commensurate with the assigned data rate of the remote unit 212. Based on the assigned data rate, the hub station 210 determines the next time T at which such bandwidth is available (i.e., not already scheduled to another remote unit transmission) over a time period which allows the remote unit 212 to exchange the desired amount of data. In this embodiment, the assigned data rate preferably remains at the highest rate or rate group possible for the remote unit to adequately transfer data.

At block 834, the hub station 210 determines if a next demand entry is in the FIFO queue waiting to be scheduled. In one embodiment, the process of Figure 10 runs continuously to handle the queue of demand entries. If another demand entry is present in the FIFO queue, the process returns to block 808 where the hub station 210 handles the demand entry, as described above. If, on the other hand, another demand entry is not present in the FIFO queue, the process terminates at block 840 or simply awaits the arrival of a next demand entry.

Figure 11 is a graphical representation of an exemplifying result of the process of scheduling one or more remote units as a function of frequency and time. Similar to the graph 800 (Figure 8), the graph 850 includes a horizontal axis representing time and a vertical axis representing the frequency spectrum. The bandwidth 842 represents the entire bandwidth available for communication by the system 200. The several numbered blocks of the graph 850 represent the bandwidth and timeslot at which a correspondingly designated remote unit is scheduled to communicate data. As an

example, the remote unit 252 is shown to be scheduled to transmit between times T9 and T10 at the indicated center frequency F and surrounding bandwidth. For the purpose of illustration, the bandwidth needed for the remote unit 212 is represented by the bandwidth 844.

As discussed above, to schedule the remote unit 212 request, the hub station 210 checks the time at which the required bandwidth 844 is available. At time T10, there may be an unscheduled timeslot 846 at its respective frequency and bandwidth. However, the timeslot 846 does not satisfy the necessary bandwidth for the remote unit 212. The hub station 210 does not schedule the remote unit 212 in the timeslot 846 because there is an insufficient bandwidth. Hence, the hub station 210 checks the next available timeslot to determine if the required bandwidth 844 for the remote unit 212 is available. At time T11, the hub station 210 finds the timeslot 212 having a sufficient bandwidth that is commensurate with the data rate of the remote unit 212. Accordingly, the hub station schedules the remote unit 212 at time T11 for a duration of a single timeslot or perhaps multiple time slots if necessary. As the hub station 210 continues to schedule transmissions, it may schedule another remote unit communication in the time slot 846.

In view of the foregoing, it will be appreciated that the invention overcomes the long-standing need for a method and system that optimizes frequency spectrum utilization among a plurality of communication stations. The system and method dynamically re-allocates the assigned frequency spectrum in response to changes in demand and frequency utilization.

Various alternative embodiments are encompassed within the scope of the invention. For example, in one embodiment, the assigned data rates are not quantized into several discrete data rates and instead each remote unit transmits the maximum data rate at which it is capable without regard to any specific data rate groups or simply using groups with a much smaller rate granularity. The invention can be applied to variety of operating environment aside from the one disclosed above with respect to Figure 1 such as terrestrial environments.

The invention may be embodied in other specific forms without departing from its spirit or essential characteristics. The described embodiment is to be considered in

all respects only illustrative and not restrictive. The scope of the invention is, therefore, indicated by the appended claims rather by the foregoing description. All changes that fall within the meaning and range of equivalency of the claims are to embraced within their scope.